

Pengelompokan Produk Pestisida Berdasarkan Pola Konsumsi Retailer Menggunakan Algoritma HAC dan K-Medoids

¹Dinda Anantya, ^{2*}Yohanni Syahra

Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

*Korespondensi: yohannisyahra@umsu.ac.id

Submit : 20 April 2026 | Diterima : 16 Mei 2026 | Terbit : 22 Mei 2026

ABSTRACT

The agricultural sector relies heavily on the effectiveness of pesticide production inputs; however, CV. Aneka Jaya Tunas Agro currently faces operational challenges due to subjective distribution decision-making. This study aims to cluster retailer consumption patterns to generate strategic insights for stock management and distribution optimization. The dataset consists of 2023–2024 sales transactions, which were cleaned into 1,488 valid records and aggregated into 134 retailer profiles. Using the Knowledge Discovery in Databases (KDD) methodology, the performance of the Hierarchical Agglomerative Clustering (HAC) algorithm with Ward's method was compared against K-Medoids (PAM) using the Silhouette Score and Davies-Bouldin Index (DBI) metrics. Test results indicate that the HAC algorithm performs better, achieving a Silhouette Score of 0.5195 and a DBI of 0.7207, compared to K-Medoids with a Silhouette Score of 0.2914 and a DBI of 1.2987. Based on the clustering results, retailers were categorized into three main segments: large retailers (10 units), medium retailers (51 units), and small retailers (73 units), each provided with specific business strategy recommendations. This study concludes that the HAC algorithm is more effective in mapping retailer consumption patterns to support data-driven decision-making.

Keywords: Clustering, Consumption Patterns, Pesticide, Hierarchical Agglomerative Clustering (HAC), K-Medoids.

ABSTRAK

Sektor pertanian sangat bergantung pada efektivitas input produksi pestisida, namun CV. Aneka Jaya Tunas Agro saat ini menghadapi tantangan operasional karena pengambilan keputusan distribusi yang cenderung subjektif. Penelitian ini bertujuan untuk mengelompokkan pola konsumsi retailer guna menghasilkan insight strategis dalam manajemen stok dan optimalisasi distribusi. Dataset yang digunakan merupakan transaksi penjualan tahun 2023–2024 yang dibersihkan menjadi 1.488 data valid dan diagregasi menjadi 134 profil retailer. Menggunakan metodologi Knowledge Discovery in Databases (KDD), dilakukan perbandingan performa antara algoritma Hierarchical Agglomerative Clustering (HAC) dengan metode Ward dan K-Medoids (PAM) menggunakan metrik Silhouette Score dan Davies-Bouldin Index (DBI). Hasil pengujian menunjukkan bahwa algoritma HAC memiliki performa lebih unggul dengan nilai Silhouette Score sebesar 0,5195 dan DBI sebesar 0,7207, dibandingkan K-Medoids yang menghasilkan Silhouette Score 0,2914 dan DBI 1,2987. Berdasarkan hasil klasterisasi, retailer dibagi menjadi tiga kategori utama, yaitu retailer besar (10 unit), retailer menengah (51 unit), dan retailer kecil (73 unit), yang masing-masing diberikan rekomendasi strategi bisnis spesifik. Penelitian ini menyimpulkan bahwa algoritma HAC lebih efektif dalam memetakan pola konsumsi retailer untuk mendukung pengambilan keputusan berbasis data (data-driven).

Kata Kunci: Klasterisasi, Pola Konsumsi, Pestisida, Hierarchical Agglomerative Clustering (HAC), K-Medoids.

PENDAHULUAN

Sektor pertanian sangat bergantung pada input produksi seperti pestisida, di mana pengecer berperan sebagai jembatan strategis antara distributor dan petani. Namun, distributor regional seperti CV. Aneka Jaya Tunas Agro masih sering menggunakan manajemen distribusi subjektif dan mengabaikan wawasan dari data historis transaksi. Pendekatan heuristik yang usang ini

memicu inefisiensi sistemik, seperti penumpukan persediaan (overstock) pada produk lambat jual dan kekosongan stok (stockout) pada produk unggulan saat musim tanam.

Untuk mengatasi inefisiensi operasional tersebut, diperlukan transformasi menuju organisasi berbasis data melalui teknik data mining dengan pendekatan unsupervised learning. Meskipun algoritma K-Means sering digunakan dalam literatur segmentasi ritel, metode ini terbukti sangat sensitif terhadap pencilan (outliers) karena menggunakan nilai rata-rata imajiner sebagai pusat klaster. Keterbatasan ini membuat K-Means kurang efektif untuk menangani data transaksi distributor yang fluktuatif dan memiliki sebaran ekstrem.

Sebagai solusi, penelitian ini membandingkan dua algoritma yang lebih kokoh, yaitu K-Medoids dan Hierarchical Agglomerative Clustering (HAC). K-Medoids menggunakan titik data nyata (medoid) untuk meredam bias pencilan, sementara HAC mampu memetakan hierarki kedekatan karakteristik pembelian secara natural tanpa menentukan jumlah klaster di awal secara deterministik. Penggunaan metode Ward's Linkage pada HAC secara matematis meminimalkan varians dalam klaster untuk menghasilkan isolasi geometri yang lebih presisi.

Penelitian ini bertujuan mengimplementasikan kerangka komparatif antara HAC dan K-Medoids untuk mengekstraksi pola konsumsi pengecer pestisida. Kualitas hasil pengelompokan divalidasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI). Melalui segmentasi yang tervalidasi ini, perusahaan diharapkan memiliki fondasi empiris untuk merestrukturisasi protokol logistik, mengoptimalkan manajemen persediaan, serta merancang strategi pemasaran yang sesuai dengan daya serap riil setiap kelas pengecer.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif komputasional dengan kerangka kerja Knowledge Discovery in Databases (KDD) yang dimodifikasi untuk menganalisis data transaksi Toko Aneka Jaya Tunas Agro periode 2023–2024. Data mentah sebanyak 1.594 rekaman melalui tahap prapemrosesan ketat, termasuk penghapusan 106 data duplikat dan penanganan missing values, menyisakan 1.488 rekaman. Inovasi utama terletak pada agregasi data dari transaksi parsial menjadi 134 profil entitas retailer tunggal untuk mengisolasi perilaku konsumsi faktual, yang kemudian dinormalisasi agar distribusi fitur menjadi homogen.

Pada tahap inti, penelitian menerapkan paradigma unsupervised learning melalui uji komparatif dua algoritma, yaitu Hierarchical Agglomerative Clustering (HAC) dan K-Medoids. HAC dijalankan dengan metode Ward's Linkage untuk meminimalkan varians dalam klaster berdasarkan Euclidean Distance. Sementara itu, K-Medoids digunakan sebagai penyeimbang yang tangguh terhadap pencilan (outliers) dengan mekanisme Partitioning Around Medoids (PAM) untuk menentukan pusat klaster dari data entitas nyata pada nilai $k=3$.

Tahap akhir difokuskan pada analisis dan evaluasi kualitas formasi klaster yang terbentuk. Validasi dilakukan menggunakan metrik Silhouette Score untuk mengukur presisi kesamaan jarak intra-klaster dibandingkan dengan jarak antar-klaster. Selain itu, Davies-Bouldin Index (DBI) diterapkan untuk menghitung tingkat pemisahan klaster, guna memastikan bahwa hasil pengelompokan pola penjualan pestisida memiliki tingkat tumpang tindih yang minimal dan struktur yang kuat.

HASIL DAN PEMBAHASAN

Pra-Pemrosesan dan Transformasi Data

Tahap pertama dalam analisis ini mengeksekusi pembersihan himpunan data transaksi operasional CV. Aneka Jaya Tunas Agro. Dari total populasi awal sebanyak 1.594 baris data mentah, algoritma data cleaning berhasil mereduksi 106 entitas duplikat serta menangani 6 baris missing values melalui imputasi. Hasilnya, dataset mengerucut menjadi 1.488 data valid. Data transaksi log harian tersebut kemudian diagregasikan ke tingkat individual, menghasilkan 134 identitas unik agregat pengecer.

Setiap entitas dari 134 pengecer tersebut kemudian direpresentasikan menggunakan empat fitur sintetik: TOTAL VOLUME, NUM_PESTICIDE_TYPES, NUM_PACKAGING_TYPES, dan rata-rata rasio belanja harian. Pada komputasi analisis deskriptif, ditemukan adanya rentang variabilitas yang masif, mulai dari volume konsumsi kelas mikro (0,50 satuan) hingga lonjakan makro di angka 2.972,00 satuan. Guna mengatasi ketimpangan rentang ini agar tidak mendominasi jarak kalkulasi (Euclidean Distance), transformasi StandardScaler (Z-score) diterapkan secara sukses menyelaraskan seluruh elemen matriks ke dalam dimensi ruang yang homogen.

Evaluasi Kualitas Model Klasterisasi

Dalam paradigma pembelajaran tanpa pengawasan (unsupervised learning), validasi model bergantung pada uji metrik internal algoritma, yang fungsinya setara dengan uji validitas statistik untuk menentukan seberapa sah titik data terkelompokkan ke dalam sebuah klaster secara presisi. Uji komparasi performa antara algoritma Hierarchical Agglomerative Clustering (HAC) dengan fungsi Ward's Linkage melawan algoritma partisi K-Medoids (PAM) dipaparkan dalam Tabel 1 berikut:

Tabel 1 Hasil Evaluasi Perbandingan Algoritma Klasterisasi

Algoritma Evaluasi Klasterisasi	Jumlah Klaster	Silhouette Score
HAC (Hierarchical Agglomerative - Ward)	3	0,5195
K-Medoids (PAM)	3	0,2914

Sumber: Data Yang Diolah (Sistem Python, 2024)

Berdasarkan Tabel 1, hasil pengujian menunjukkan secara kuantitatif bahwa algoritma HAC mengungguli K-Medoids secara mutlak. Algoritma HAC mencetak Silhouette Score sebesar 0,5195, yang mengindikasikan tingkat homogenitas (kemiripan anggota) di dalam klaster tergolong padat dan valid. Sebaliknya, K-Medoids merosot pada skor 0,2914. Parameter ini didukung lebih jauh oleh uji Davies-Bouldin Index (DBI) di mana HAC berhasil menekan tingkat overlap antar klaster ke titik serendah 0,7207, jauh lebih superior ketimbang pemisahan K-Medoids yang tercampur baur pada DBI 1,2987.

Distribusi Hasil Klasterisasi

Mempertimbangkan supremasi metrik dari algoritma HAC, maka model ini dipilih sebagai instrumen utama pemetaan. Pemotongan dendrogram di ketinggian ambang batas hierarki (sekitar 16,34) memecah 134 identitas pengecer distributor secara natural ke dalam 3 ekuilibrium klaster (K=3) dengan distribusi berikut:

Tabel 2 Distribusi Demografi Pengecer (Segmentasi HAC)

Identitas Segmentasi	Populasi (Unit)	Persentase
Klaster 1 (Retailer Besar)	10 pengecer	7,5%
Klaster 0 (Retailer Menengah)	51 pengecer	38,1%
Klaster 2 (Retailer Kecil)	73 pengecer	54,5%

Pembahasan

Hasil penelitian ini membuktikan bahwa penambahan data algoritmik berdampak positif dan signifikan dalam mengekstraksi pengetahuan yang tersembunyi dari tumpukan histori transaksi penjualan. Berdasarkan komparasi evaluasi, algoritma Hierarchical Agglomerative Clustering (HAC) terbukti lebih sah dan presisi dibandingkan K-Medoids dalam membedah lanskap ritel komoditas pestisida. Ketertinggalan algoritma K-Medoids dapat dikorelasikan pada sifat biologis dari pertumbuhan industri ritel itu sendiri. Metode berbasis titik tengah penugasan partisi (seperti K-Medoids/K-Means) selalu memaksa populasi data untuk membentuk bola sferis yang melingkar sempurna. Pada kenyataannya, profil toko saprotan bertumbuh secara asimetris melintasi spektrum kontinuitas—dari entitas eceran mikro merangkak naik menjadi agen partai besar. Arsitektur HAC yang bekerja melalui penautan penggabungan akar rumput (bottom-up) sangat fleksibel menelusuri variabilitas tajam ini tanpa memerlukan ilusi pusat gravitasi statis.

Intinya, wawasan klasifikasi perilaku konsumsi ini harus menjadi landasan bagi CV. Aneka Jaya Tunas Agro dalam mengeksekusi strategi berbasis bukti empiris (data-driven strategy), menggantikan pola kerja lama yang diasumsikan secara subjektif. Apabila rekomendasi analitik pada Tabel 2 ini diimplementasikan, maka hubungan manajemen logistik antara distributor dan konsumen perantaranya akan menjadi sangat harmonis dan berpotensi memotong biaya inefisiensi yang besar. Penjelasan teknis dari masing-masing segmen dijabarkan sebagai

berikut:

1. Strategi untuk Klaster 1 (Retailer Besar / Prioritas): Berjumlah 10 agen dengan daya beli dan perputaran uang masif. Keberhasilan perusahaan mengamankan kepuasan pelanggan pada segmen ini akan menjaga stabilitas fundamental distributor. Strategi yang paling disarankan adalah penerapan tata kelola pasokan Vendor Managed Inventory (VMI). Distributor harus proaktif mengontrol pemesanan mereka dalam kapasitas curah besar (drum) serta mengikat mereka melalui skema rabat loyalis (volume discount tiering).
2. Strategi untuk Klaster 0 (Retailer Menengah / Agresif): Memiliki kekuatan 51 unit pengecer. Peran mereka penting sebagai titik perputaran roda finansial yang stabil. Untuk meningkatkan utilitas pesanan mereka, distributor harus mengeksekusi promosi bersubsidi silang (cross-selling atau bundling). Obat pembasmi hama (insektisida) yang memiliki tingkat serapan tinggi dapat dikemas satu harga (bundel) berdampingan dengan komoditas herbisida spesifik yang pergerakannya melambat di gudang distributor, sehingga mengatasi ancaman penumpukan overstock.
3. Strategi untuk Klaster 2 (Retailer Kecil / Akar Rumput): Dengan komposisi padat berjumlah 73 unit (54,5% keseluruhan populasi). Mereka menghadapi keterbatasan perputaran arus kas harian. Distributor tidak boleh memaksa pengiriman pesanan dalam jumlah besar dengan MOQ tinggi yang dapat mencekik kelangsungan bisnis segmen ini. Taktik logistik perlu direstrukturisasi dengan frekuensi kedatangan barang yang lebih cepat (pengiriman repetitif), memfokuskan armada untuk mendistribusikan inventaris pestisida berukuran mikro seperti sachet dan botol berukuran 100 mililiter hingga 250 mililiter. Dukungan tambahan berbasis layanan, seperti menyediakan material pemasaran (spanduk edukasi bahan aktif), akan memungkinkan pengecer kelas mikro ini menjangkau langsung populasi petani perseorangan.

Strategi ini beresonansi erat dengan kemampuan penyesuaian layanan terhadap klasifikasi kebutuhan pelanggan. Memahami demografi data transaksi merupakan pondasi esensial untuk mendongkrak keuntungan yang sejalan dengan peningkatan tingkat retensi pengecer.

KESIMPULAN

Investigasi penambangan pengetahuan empiris pada rantai distribusi ini menegaskan urgensi adopsi kecerdasan analitik demi meruntuhkan kultur heuristik serta operasional pasif di CV. Aneka Jaya Tunas Agro. Proses uji komparasi data memvalidasi secara mutlak bahwa kerangka model algoritma Hierarchical Agglomerative Clustering (HAC) ditopang metode ukur Ward's Linkage sanggup menghasilkan struktur klasterisasi spasial yang kokoh, dibuktikan oleh pencapaian ekuilibrium Silhouette Score pada angka 0,5195 dan supresi tumpang tindih objek (DBI) pada nilai ideal 0,7207, meninggalkan performa K-Medoids yang jauh lebih rendah di segala sisi metrik evaluasi. Konfigurasi algoritmik ini mengekstrak dan memilah entitas pasar menjadi tiga strata definitif yang langsung dapat diterjemahkan menjadi protokol intervensi bisnis. Distributor dapat mengaplikasikan tata kelola Vendor Managed Inventory (VMI) bervolume berat pada klaster 1 (pengecer besar), rekayasa pemasaran subsidi silang portofolio produk di klaster 0 (pengecer menengah), dan adaptasi rute logistik asinkron berfrekuensi tinggi dengan distribusi kemasan mikro di klaster 2 (pengecer kecil).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Rektor Universitas Muhammadiyah Sumatera Utara (UMSU) Prof. Dr. Agussani, M.AP., Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) Dr. Al-Khowarizmi, S.Kom., M.Kom., serta Ketua Program Studi Teknologi Informasi Ibu Fatma Sari Hutagalung, S.Kom., M.Kom., atas fasilitas, dukungan, dan kesempatan yang telah diberikan. Ucapan terima kasih juga penulis sampaikan kepada Ibu Yohanni Syahra, S.Si., M.Kom., selaku dosen pembimbing, atas bimbingan, arahan, dan kesabaran selama proses penyusunan penelitian ini. Selain itu, penulis turut mengucapkan terima kasih kepada kedua orang tua tercinta serta rekan-rekan, yaitu Aida Fadhila, Dwi Nisyatul Wardah, dan Farhan Fanalty, atas doa, dukungan, dan motivasi yang senantiasa diberikan hingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

- Agustin, E. W., et al. (2025). Optimization of Customer Segmentation in the Retail Industry Using the K-Medoid Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 766-775.
- Ahmad, S. F., et al. (2023). Data Mining Methods and Obstacles: A Comprehensive Analysis. *Artificial Intelligence Review*, 56(1).
- Chilipweli, P. M., Nyanza, E. C., & Ngowi, A. V. (2025). Patterns, Practices, and Socio-Environmental Dynamics of Pesticide Use in the Horticultural Value Chain: Insights from Smallholder Farmers and Agro-Input Sellers. *Agriculture*, 15(2), 1-24.
- Dhulipala, L., et al. (2022). Hierarchical Agglomerative Graph Clustering in Poly-Logarithmic Depth. *Advances in Neural Information Processing Systems*, 35(NeurlIPS).
- Farahnakian, F., et al. (2023). A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behavior. *Remote Sensing*, 15(6), 1-34.
- Garcia-Morato, E., et al. (2024). A general framework for distributed approximate similarity search with arbitrary distances. *arXiv preprint*, arXiv:2405.13795.
- John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, 2(4), 809-823.
- Kajornkasirat, S., et al. (2025). Customer segmentation using association rule mining on retail transaction data. *IAES International Journal of Artificial Intelligence*, 14(3), 1919-1929.
- Luh, N., et al. (2022). Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara). *Jurnal Sistim Informasi dan Teknologi*, 16(2), 105-112.
- Mathelin, A. de., et al. (2025). OneBatchPAM: A Fast and Frugal K-Medoids Algorithm. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15), 16172-16180.
- Noval, N., et al. (2025). Implementation of Data Mining to Categorize Daily Sales Transaction Data Using K-Means. *Journal of Information Systems*, 6(1), 1-12.
- Oti, E. U., & Olusola, M. O. (2024). Overview of Agglomerative Hierarchical Clustering Methods. *British Journal of Computer, Networking and Information Technology*, 7(2), 14-23.
- Pireddu, A., et al. (2024). A Review of Data Mining Strategies by Data Type, with a Focus on Construction Processes and Health and Safety Management. *International Journal of Environmental Research and Public Health*, 21(7).
- Rizky, F., et al. (2025). Implementation of Data Mining to Categorize Daily Sales Transaction Data Using K-Means. *Jurnal Sistim Informasi dan Teknologi*, 6(1), 2160-2169.
- Shu, Y., & Ye, X. (2023). Knowledge Discovery in Databases and Data Mining Applications in Logistics and Supply Chain Management. *Information Systems Research*, 22(3), 114-129.
- Syahra, Y., et al. (2025). Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry. *PeerJ Computer Science*, 9(3), 1120-1131.
- Tiwari, M., et al. (2023). BanditPAM++: Faster k-medoids Clustering. *Advances in Neural Information Processing Systems*, 36(NeurlIPS).
- Triantara, I. K. Y., Raharja, M. A., & Sarasvananda, I. B. G. (2025). Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering. *JNATIA*, 3(4), 1-10.
- Tsiu, S. V., et al. (2025). Applications and Competitive Advantages of Data Mining and Business Intelligence in SMEs Performance: A Systematic Review. *Businesses*, 5(2), 22.